Remarks

Claims 1-13 are pending in the application. Claims 1-13 are rejected. Claims 1, 4, 5, 8 and 9 are amended. All rejections are respectfully traversed.

Claims 1-13 are rejected under 35 U.S.C. 101 because the claims are not directed toward statutory subject matter. The claims have been amended to define a computer implemented process.

The invention provides a method for determining similarities of interpretation between portions of multimedia (videos) at a very high level, e.g., similar action in an adventure movie, scoring opportunities in a sports video, romantic activity in a gothic movie, fright in a horror movie, humor in a comedy movie, and so forth. The term 'high-level' is used because the similarity considers a sequence of semantic events extended over a relatively long time period. Low-level similarities would consider color in individual frames taken in only a fraction of a second.

The similarity is determined by comparing ordered content entities in directed acyclic graphs (DAGs) of the video. The reason that the similarity is high-level is that the comparison according to the invention orders several content entities in DAGs, and not just a single content entity, such a "shot" or a frame.

For example, a *high-level* "scoring opportunity" could include a content entity (a shot) of two soccer players making a break-away with the ball,

followed by a shot of the player firing the ball at the goal, followed by a shot of the ball being deflected. Here the order is temporal, and different camera shots can be linked. The break-away may be a wide angle shot of several players, the shot on goal may be a close up of one player, and the deflection a medium distance shot of the defending goalie.

Romantic activity in a movie might be a meaningful look, followed by a hug and a kiss. Other high-level interpretations of portions of videos of different genres can easily be defined.

It is the way that that these various shots are put together (ordred) that defines the high-level activity and interpretation, and this is the interpretation that is captured by the ordered content entities in the claimed DAGs. For example, the comparison according to the invention can detect similar "scoring" opportunities, even though the players are different in the different shots, and the timing may be quite different.

It is the stringing together of several, perhaps dissimilar, content entities (shots) in an ordered manner in multiple DAGs to encode and compare high-level interpretations of the content that makes the invention novel.

Claims 1-13 are rejected under 35 U.S.C. 103(a) as being unpatentable over Yeo et al., U.S. Patent No. 5,821,945 (Yeo).

Yeo also describes similarities in videos. However, the similarities in Yeo are at a very low level.

Directed graph representation of shots

There is one level of hierarchy, i.e. H=1. F_0 partitions $\{s_i\}$ into $V_{0,1}$, $V_{0,2}$, . . . , such that nodes in each $V_{0,i}$ are sufficiently similar, according to some similarity measured in terms of low level vision indices such as colors, shapes, etc.

Yeo also describes similarities between shots.

In this case, shots that are similar to each other are clustered together. Relations between clusters are governed by temporal ordering of shots within the two clusters. A simple example would be a scene of conversation between two persons; the camera alternates between shots of each person. The graph G_0 consists of two nodes $V_{0,1}$ and $V_{0,2}$; ($V_{0,1}$, $V_{0,1}$) and ($V_{0,2}$, $V_{0,2}$) are both members of E_0

and

Similarity of shots

Low level vision analyses operated on video frames achieve reasonably good results for the measurement of similarity (or dissimilarity) of different shots. Similarity measures based on image attributes such as color, spatial correlation and shape can distinguish different shots to a significant degree, even when operated on much reduced images as the DC images. Both color and simple shape information are used to measure similarity of the shots.

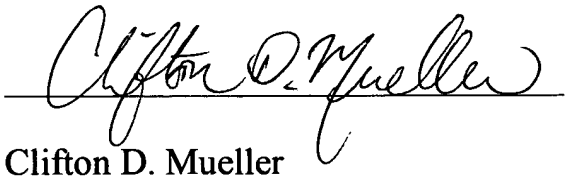
The example described by Yeo for the latter case is where he can distinguish Mr. A and Ms. B in different shots, see column 8.

However, Yeo never orders different shots in a related sequence of shots using a DAG to provide a high-level interpretation of what is going on at a semantic level.

Obviously, this type of similarity does not reach the higher-level interpretations of what is claimed. Other low-level features described by Yeo to express similarities of shots include color histograms, pixel luminance, and shape, see Figure 11. However, Yeo does not order shots in his transition graph and compare the graphs. Yeo does not describe the comparison of ordered content entities (in Yeo, "shots") in a plurality of the directed acyclic graphs (in Yeo, "transition graphs') to determine a similarity of interpretations of the multimedia content. Yeo does not order content entities, spatially or temporally, according to intensity attributes or direction attributes.

It is believed that this application is now in condition for allowance. A notice to this effect is respectfully requested. Should further questions arise concerning this application, the Examiner is invited to call Applicants' agent at the number listed below. Please charge any shortage in fees due in connection with the filing of this paper to Deposit Account 50-0749.

Respectfully submitted,
Mitsubishi Electric Research Laboratories, Inc.

By 

Clifton D. Mueller
Agent for the Assignee
Reg. No. 57,836

201 Broadway, 8th Floor
Cambridge, MA 02139
Telephone: (617) 621-7517
Customer No. 022199